

Entrega Etapa Ideação

Equipe 28 – Noble Eagles

Processo

O processo de ideação, feito utilizando o Excalidraw, nos apoiou no processo de análise das ideias e compreensão das informações disponibilizadas e seu papel na resolução do problema definido na etapa de Imersão.

Descritivo da Etapa anterior

Problema:

À partir dos principais problemas reportados pelo SINE, identificamos que, apesar de haver um grande número de profissionais e vagas, há uma dificuldade em promover o encontro entre empregador e trabalhador.

Portanto, autaremos para melhorar a capacidade do SINE de indicar vagas para os trabalhadores e profissionais para os empregadores garantindo maior assertividade nas entrevistas. Ou seja, que o trabalhador tenha que se deslocar para menos entrevistas e que o empregador dedique menos esforço com o recrutamento.

Hipóteses:

Nesta etapa mapeamos algumas hipóteses iniciais que podem nos apoiar na solução do problema:

- É possível melhorar a probabilidade de contratação estabelecendo relacionamentos entre região, CBO, descritivo das ocupações ofertadas e capacidades e habilidades do trabalhador utilizando as bases do CAGED, CBO, SINE Trabalhadores e SINE Vagas
- Tanto trabalhadores como empregadores podem se beneficiar de um modelo de sugestões sobre habilidades necessárias e disponíveis no mercado

Riscos

- Há imprecisão no descritivo das habilidades necessárias para o desempenho das atividades requeridas pelo empregador
- Há imprecisão no descritivo das capacidades e habilidades do trabalhador
- A exigência da vaga vai além ou fica aquém do trabalho a ser realizado, causando problemas no processo de recrutamento
- Oferta e demanda não estão na mesma área geográfica

Entrega

Buscaremos uma forma de construir um algoritmo de match making para aumentar a probabilidade de trabalhador e empregador se encontrarem por meio do SINE

Resumo das ideias

Ideias adotadas

- Utilização do Azure SQL para carga e consolidação dos arquivos do SINE Empregos, SINE Trabalhador, CAGED e CBO
- Aplicação de regressão logística por área para a predição da probabilidade de sucesso
- Utilização da base do CAGED como base de 'empregados', utilizando CBO, UF, deficiência (S/N) e escolaridade para treinar o modelo
- Utilização da variável "admitido/demitido" como variável dependente para treino do modelo
- Utilização do Azure Machine Learning para lidar com o volume de informações (que excede em muito a capacidade de processamento local das máquinas do time)

Ideias abandonadas

- Dados informações de seguro desemprego: entendemos que, por serem dados agregados, teriam pouco potencial de contribuição para a criação do modelo preditivo
- Base da RAIS: entendemos que o CAGED de 2019 possuía as informações necessárias para a modelagem
- BASE CAGED 2020: Poucas informações e dados fortemente impactados pela Covid-19
- Utilização das bases do SINE tanto de vagas como de trabalhadores para treinar um modelo de machine learning, pois, devido à presença de informações sensíveis, o volume de dados disponibilizados é insuficiente

Desenho geral da solução adotada

Com base nesta etapa iniciamos o desenho da solução a ser prototipada:

